Stochastic Defense Against Complex Grid Attacks

Daniel Bienstock and Mauro Escobar

***Abstract*—We describe stochastic defense mechanisms designed to detect sophisticated grid attacks involving both physical actions (including load modification) and sensor output alteration. The initial attacks are undetectable under a full ac power-flow model even assuming ubiquitous sensor placement, while hiding large line overloads. The defensive techniques apply network control actions that change voltages in a random fashion and additionally introduce (random) low-rank corrections to covariance matrices.**

***Index Terms*—Cyber-physical power grid attacks, security.**

I. Introduction

RECENT events and research efforts have highlighted the potential for powerful coordinated attacks on power grids that combine disruption or modification of sensor data with physical actions. Such attacks may succeed in hiding from operators' undesirable system conditions, long enough that the physical damage or automatic shutdown of equipment takes place, an undesirable and potentially risky outcome.

We propose techniques to be deployed when a high-fidelity attack on a power grid is suspected, with details unknown. These techniques involve the following two ideas:

1) using network resources to randomly change power flow quantities, especially voltages and, in particular
2) changing the covariance structure of, e.g., voltages in a manner unpredictable by the attacker. The specific version of this idea that we analyze introduces a low-rank adjustment to the covariance of phase angles.

Attacks of concern are those whose data component is designed to pass a stringent test, namely, that the falsified data satisfy the full ac power flow equations (see [1]–[3]) at every bus and line. The data attack is coordinated with a physical attack that causes a dangerous system condition, e.g., a line overload. The data modification hides this overload, with the result that sensor data received by operators are both unimpeachable and portray safe system operation. Alternatively, a data-only attack may portray a false situation with the goal of causing an improper control action. Putting aside the actual feasibility of such attacks, the computational challenge is significant. Prior work (e.g., [4]) has already addressed this problem. As motivation for our focus on defense, we develop a new optimization formulation that quickly computes successful attacks on large transmission systems with thousands of buses.

A. Prior Work Using Static State Estimation

The possibility of cyber- or cyber-physical attacks on power grids has yielded mathematical work designed to detect and reconstruct such attacks (see [4]–[25]).

The starting point of this article is the currently used "State Estimation" procedure wherein sensor readings are used together with a linearized model of power flow in order to estimate other system parameters. In its simplest form, this procedure uses the linearized, or dc power flow model [1]–[3] (also see Section II)

$$B\theta \;=\; P^g - P^d \tag{1}$$

where B is the bus susceptance matrix, θ is the vector of phase angles, and P^g and P^d are the vectors of active power generation and load, respectively. Sensors, which may not be ubiquitous, report phase angles, and statistical estimation procedures can be used to recover missing readings as well as other operational data. If the estimated data fail to satisfy (1), an anomalous condition may be construed. As discussed in the abovementioned works, an attacker that is able to modify sensor output may be able to alter the true phase angles θ through a perturbation δ in the null space of B; the vector $\theta' \doteq \theta + \delta$ is, thus, consistent with the (1). The resulting attack is, thus, considered *undetectable* as per the state estimation criterion. Attacks may include a physical component that modifies the underlying network, thus providing an additional challenge to a system operator. As shown in [5] such attacks may be sparse (i.e., δ has very small support); computation of an optimally sparse attack is considered in [6].

When the defender does not have unhindered access to sensors, or if, e.g., the result of the attack is that sensors stop reporting, sophisticated techniques may still be brought to bear in order to identify, for example, the topology modification (see [7]–[10]).

Some the abovementioned work relies on dc-based state estimation; the model in [8] uses the ac power flows model (also see [12]). Attacks that modify admittances are considered

Manuscript received October 5, 2019; revised October 9, 2019; accepted October 13, 2019. Date of publication October 29, 2019; date of current version June 12, 2020. This work was supported in part by the Duke of Edinburgh's Award GMLC77 under Grant HDTRA1-13-1-0021, in part by the Office of Naval Research under Grant N00014-16-1-2889, in part by the Defense Advanced Research Projects Agency (DARPA), in part by the Rapid Attack Detection, Isolation, and Characterization Systems (RADICS) Award, and in part by the DARPA Lagrange Award. Recommended by Associate Editor P. Barooah. *(Corresponding author: Daniel Bienstock.)*

The authors are with Columbia University, New York, NY 10027 USA (e-mail: dano@columbia.edu; me2533@columbia.edu).

Digital Object Identifier 10.1109/TCNS.2019.2949908

in [13]. Valenzuela *et al.* [21] performed principal component analysis (PCA) on the covariance of power flows to discover anomalies by inspecting changes in the smaller eigenvalue modes, also see [23]–[28].

Zhang *et al.* [4] (also see [11]) studied ac-undetectable cyber-physical attacks. Zhang and Sankar [4] solved an optimization problem to compute a line to be disconnected by the attacker, and data modifications, so that a resulting line overload is hidden; attacks on the IEEE 24-bus model are reported.

B. Prior Work Relying on Dynamics

A cyber-physical attack would likely engender transients and dynamic behavior, suggesting the use of detection tools that focus on the fidelity of sensor signals in a dynamic regime. This is the approach taken in, e.g., [29] and [30], which consider abstract linearized models of a system under cyber-physical attack. Under appropriate assumptions, the detection algorithms described therein are guaranteed to detect attacks. In the power setting, the approach in [29] or [30] amounts to a linearization of the swing equation [2], [3] at generator buses. At present, the issue of how to model frequency dynamics at load buses is an open research question; how to model voltage magnitude dynamics also presents difficulties. However, techniques such as those in [29] or [30] could be deployed to rapidly highlight the possibility of an unexplained event.

C. Semantics

There is a large amount of literature in the power engineering, control, and signal processing communities devoted to “cyber-physical” attacks, very broadly defined. In this section, we aim to pin down some of the semantics.

First, insofar as there is an *attacker*, we will use the term *defender* to describe the entity that seeks to protect the network. This defender can be taken as synonymous with a control center for the power system. In this article, the actions of the defender will primarily seek to understand the nature of the attack, i.e., to detect the attack.

Another issue concerns the timing of the attack. See Sections III–V for details concerning the nature of the attack. We assume that the attacker has gathered comprehensive data on the system. The actual attack begins at a certain point in time; at that point the attacker carries out the physical component of the attack, if any. From then on, the attacker causes attacked sensors to output dynamically falsified data.

We assume that an attack becomes suspected, at some point after the attack begins, as a result of monitoring techniques, e.g., those in [29] or [30]. At that point, the techniques discussed in this article are deployed.

The defensive techniques that we describe may require minutes of elapsed time before the attack is detected. An attack that creates a massive line overload or voltage drop, may, thus, go undetected before secondary events such as line trips or system-wide voltage collapse take place. However, the defensive techniques may succeed under less severe (but still risky) circumstances.

II. Notation

We represent ac power flows using the polar representation. The voltage at a bus k is of the form $V_k = |V_k|e^{j\theta_k}$ where $j = \sqrt{-1}$. A line km is described by using the standard “π” model that includes series impedance, line charging, and transformer attributes (see, e.g., [2] and [3]). Under this model, the complex current injected into line km at bus k and m, respectively, are given by the formula

$$\begin{pmatrix} I_{km} \\ I_{mk} \end{pmatrix} = Y_{km} \begin{pmatrix} V_k \\ V_m \end{pmatrix} \tag{2}$$

where Y_{km} is the branch admittance matrix for line km; the complex power injected into line km at k equals $p_{km} + jq_{km} = V_k I^*_{km}$. Here, $p_{km} = p_{km}(V_k, V_m)$ and $q_{km} = q_{km}(V_k, V_m)$ are real-valued quadratic functions of the voltages at k and m, which can be summarized in the form

$$p_{km} + jq_{km} = S_{km}(|V_k|, |V_m|, \theta_k, \theta_m). \tag{3}$$

The complex power flow and angle limits on a line km are denoted by $S^{\max}_{km}$ and $\theta^{\max}_{km}$, respectively, the voltage limits at a bus k are given by $V^{\min}_k$ and $V^{\max}_k$, and the active and reactive limits at a generator bus k are indicated by $P^{g,\min}_k$, $P^{g,\max}_k$ and $Q^{g,\min}_k$, $Q^{g,\max}_k$, respectively.

Given a bus k, we denote by $\delta(k)$ the set of all lines of the form km. $\mathcal{N}$ is the set of buses (we write $n = |\mathcal{N}|$) and $\mathcal{G}$ is the set of generator buses;[1] given a set of buses S we denote by ∂S (the boundary of S) the subset of buses of S that are incident with a line with an end not in S.

We model automatic generation control (AGC) as follows. There is a selected subset of generators $\mathcal{R} \subseteq \mathcal{G}$ (the participating generators) and parameters $\alpha_k \geq 0$ for $k \in \mathcal{R}$ (the participation factors) with $\sum_{k\in\mathcal{R}} \alpha_k = 1$. If aggregate net active power generation changes by some value Δ, with generator $k \in \mathcal{R}$ changing its output by $\alpha_k\Delta$.

The susceptance matrix B of the dc power flow model (1) is defined by $B_{kk} = \sum_{km\in\delta(k)} 1/x_{km}$ for any bus k, $B_{km} = -1/x_{km}$ for any line km, and $B_{km} = 0$ otherwise; where $x_{km} > 0$ is the reactance of line km. When the underlying network is connected, the solution to system (1) has an important attribute, namely that it has one degree of freedom. Any arbitrary bus (the *reference* bus) can be selected, and its phase angle set to zero—with this proviso, system (1) has a unique solution.

A. Static AC State Estimation

The counterpart to traditional dc power flow state estimation is ac state estimation, which analyzes the residuals to ac power flow laws using sensor outputs. The control center receives, from a sensor located at a bus k, a reading V^{R}_k for the complex-valued voltage at k. Likewise, for any line km, the control center receives readings I^{R}_{km} and I^{R}_{mk} for the complex-valued current injections at k and m, assuming appropriate sensors. Here, the “R” superscript is used to stress the reported nature of the readings, as opposed to true (exact) physical values. All outputs are

[1] For simplicity, we assume at most one generator per bus.

time-stamped. Sensors also output frequency estimations, but those are not used in the consistency conditions given next.

There are two basic consistency requirements that the reported data such satisfy. First, for any line km the quantities $V_k^{\mathrm{R}}, V_m^{\mathrm{R}}, I_{km}^{\mathrm{R}}$, and I_{mk}^{R} satisfy *current-voltage consistency*

$$\begin{pmatrix} I_{km}^{\mathrm{R}} \\ I_{mk}^{\mathrm{R}} \end{pmatrix} = Y_{km} \begin{pmatrix} V_k^{\mathrm{R}} \\ V_m^{\mathrm{R}} \end{pmatrix} \tag{4}$$

for any line km, i.e., (2).[2] A second condition that the outputs must satisfy is that for any bus k we have *power-injection consistency*, that is

$$\sum_{km \in \delta(k)} V_k^{\mathrm{R}} I_{km}^{\mathrm{R}*} = \text{net injection at } k, \quad \forall k \tag{5}$$

we assume that the defender has an accurate estimate of the right-hand side of (5).

Under normal operation, equipment limits should not be exceeded, that is to say for any line km it should be the case that $|S_{km}| \leq S_{km}^{\max}$ (and similarly for mk), for any bus k it should hold that $V_k^{\min} \leq V_k \leq V_k^{\max}$ and for any generator bus k, the total (active and reactive) generation should be in the ranges $[P_k^{g,\min},\ P_k^{g,\max}]$ and $[Q_k^{g,\min},\ Q_k^{g,\max}]$, respectively. A violation of any of these conditions, if detected, would raise an alarm.

B. Organization of the Article

In Sections III–V, we discuss a specific version of the attacks considered in this article. Our defensive techniques are presented in Sections VI–VII. A discussion of the computation of the attacks as optimization problems is given in the Appendix.

III. Structure of Attacks

To motivate the development of our defensive techniques, in this section, we describe a type of complex attack that should prove difficult to identify. Our assumptions are as follows.

(s.1) The attack takes place during a period of slowly changing ambient conditions (especially loads). The attacker has complete knowledge of the network and average ambient conditions.

(s.2) Each bus k there is a sensor measuring voltage at k and current at each line $km \in \delta(k)$.

(s.3) The attacker has selected a subset $\mathcal{A}$ of buses, the focus of the attack. We assume $\mathcal{G} \cap \mathcal{A} = \emptyset$ (no generator buses within $\mathcal{A}$).

(s.4) The attacker's actions (physical and data modifications) are taken within $\mathcal{A}$.

The actual attack we consider is made up of the following two phases.

1) An *initial* phase, where the physical component of the attack takes place; it is assumed that this phase is very rapid. We will use the terms "initial phase" and "initial attack" interchangeably. In Sections IV and VIII, we will provide details on a precise mathematical task performed by the attacker in order to compute the physical component of the attack and a corresponding set of static falsified sensor readings (within $\mathcal{A}$) that guarantee ac undetectability of the attack in a static sense. We will refer to the pair consisting of the physical actions and the static readings as a *static attack*.
2) A *follow-up* phase, where at each point in time following the initial attack, sensors within $\mathcal{A}$ report perturbations to the sensor readings computed as part of the static attack (see Section V).

Discussion: Immediately following any physical modification to a system, we can expect a change in voltages (magnitudes and phase angles) and even to system frequency, the latter especially when net loads are changed. More properly, system dynamics will undergo a change, and generator frequency dynamics can be modeled using the so-called swing equation, but a precise understanding of overall grid dynamics remains a challenging computational task. See [31] for a survey of the current state-of-the art.

The attributes of the (static) attack paradigm as given in Section IV guarantees static ac state estimation undetectability at the onset of the attack and accurate representation of AGC responses, if any; the data perturbations applied in the follow-up phase will guarantee approximate static undetectability even as ambient conditions change. Assuming that the initial phase has caused an equipment overload, a delay in undetectability may lead to secondary (undesirable) events.

IV. Static Attacks

In this section, we discuss a specific version of the static attacks introduced previously. We refer the reader to [4] or [13] for additional related discussions and algorithms. In the Appendix, we provide a nonlinear, single-phase optimization problem to compute the desired physical modifications and data distortion. We show that the numerical solution to this problem scales well to systems with thousands of buses, with running times in the tens of seconds or less on a standard computer.

In (s.1)–(s.4), we outlined the conditions required for the static attack. Template IV.1 (similar to one in [4]) presents these conditions in a more explicit form. The discussion following the template will provide specific details.

Template IV.1: AC-undetectable static attack.

a) The attacker has selected a target line uv within $\mathcal{A}$ that will be overloaded.

b) The attacker's physical actions are of several types: The attacker may modify *loads* at buses in $\mathcal{A}$, may disconnect lines with both ends in $\mathcal{A}$ and may alter admittances of lines with both ends in $\mathcal{A}$.

c) For any bus $k \in \mathcal{A}$, the attacker may modify data provided by a sensor located at k.

d) The data received by the control center satisfies voltage-current and power-injection consistency and shows all system limits being satisfied, while in actuality line uv is overloaded.

e) When the attack includes load changes, secondary response (i.e., AGC response) is taken into account by the attacker.

[2] We note that the determination of the matrix Y_{km} is itself a part of the state estimation task; a discussion is beyond the scope of this article.

Note that we allow loads to be modified, but not generation.

Next we expand on the conditions provided by the template. In what follows, *true* data will be the true physical data, given by the (voltage, current) pair of vectors $(V^{\mathrm{T}}, I^{\mathrm{T}})$ In contrast, *reported* data are that which is actually received by the control center; as before given by $(V^{\mathrm{R}}, I^{\mathrm{R}})$. First, we will require that the reported data satisfy current-voltage consistency, i.e., condition (4). This condition will be enforced in the computation given in the following in an indirect fashion (also see [13] for a different use of this requirement).

1) On a bus $k \notin \mathcal{A}$, the true and reported data agree (no data modification outside $\mathcal{A}$, by definition).
2) At a bus $k \in \partial\mathcal{A}$, the attacker is constrained by the condition $V_k^{\mathrm{R}} = V_k^{\mathrm{T}}$. This condition is applied to avoid attack detection, given (a) and the second equation in (4) applied to a line km where $m \notin \mathcal{A}$.
3) On buses $k \in \mathcal{A} - \partial\mathcal{A}$, we may have $V_k^{\mathrm{R}} \neq V_k^{\mathrm{T}}$ and on lines with at least one end in $\mathcal{A} - \partial\mathcal{A}$, the true and reported currents may also differ.
4) Outside of $\mathcal{A}$, the reported voltages and currents satisfy power-injection consistency together with the preattack net-injections. Consider a bus k. The quantity $\sum_{km \in \delta(k)} V_k^{\mathrm{R}} I_{km}^{\mathrm{R}*}$ equals the power injected into the system at bus k, as per the reported data. If $k \notin \mathcal{A}$ by definition (of reported and true data) this sum equals $\sum_{km \in \delta(k)} V_k^{\mathrm{T}} I_{km}^{\mathrm{T}*}$, which is the true net power injected at bus k; power-injection consistency implies that this quantity will be unchanged as a result of the attack. On the other hand if $k \in \mathcal{A}$ the sum may differ from the true injection at k.

(s.5) If the attack causes a net change in the sum of loads, the resulting AGC-mandated change in generator output must be taken into account.

In the Appendix, we will provide an optimization problem whose solution attains the abovementioned goals, as well as experiments showing that its solution is efficiently attained, yielding successful attacks on large systems.

V. Follow-Up Phase

Previously, we described actions to be taken by an attacker in order to initiate an undetectable attack. In order for the attack to be truly successful, it must remain undetected or at least unreconstructed for a sufficiently long period of time, possibly on the order of minutes. This presents a challenge to the attacker, as the modified sensor data must paint a falsified picture, yet such sensor data cannot be constant and more generally must follow a realistic stochastic distribution.

Let $t = 0$ denote the time at which the initial attack is completed. In analogy to our notation for the initial problem, at time $t > 0$, we denote by $V_k^{\mathrm{R}}(t)$ and $V_k^{\mathrm{T}}(t)$, the reported and true voltages at t and similarly with currents. Reported data for $\mathcal{A}$ will be created by the attacker aiming to approximately satisfy current–voltage and power-injection consistency [see (4) and (5)].

In addition, in this article, we assume that the attack is perpetrated when ambient conditions (in particular loads) are, on average, constant. Let us denote by $V_k^{\mathrm{R}}(0)$ the voltage at a bus k computed by the initial attack, i.e.,

$$V_k^{\mathrm{R}}(0) \ \doteq\ |V_k^{\mathrm{R}}| e^{j\theta_k^{\mathrm{R}}}$$

and likewise define the current $I_{km}^{R}(0)$ on line km. The statement that ambient conditions are approximately constant, post attack, can be informally rephrased as

$$V_k^{\mathrm{R}}(t) \ \approx\ V_k^{\mathrm{R}}(0) \, \forall\, k, \text{ and } I_{km}^{\mathrm{R}}(t) \ \approx\ I_{km}^{\mathrm{R}}(0) \, \forall\, km. \qquad (6)$$

If ambient conditions are approximately constant (6) will hold (statistically) for any bus km not in the attacked zone $\mathcal{A}$ but are otherwise a requirement for the attacker.

Two types of attack have been used in the literature. First, the "noisy data" attack in our setting works as follows:

Template V.1: Noisy data attack.

At time $t > 0$, the attacker reports at each bus $k \in \mathcal{A}$ a voltage $V_k^{\mathrm{R}}(t) \ = \ V_k^{\mathrm{R}}(0) + \boldsymbol{\nu_k(t)}$.

Here, $\boldsymbol{\nu_k(t)}$ is a random[a] value drawn from a zero mean distribution with small standard deviation. (See discussion in the following on the meaning of "small"). Likewise the attacker reports for each line km with both ends in $\mathcal{A}$, currents

$$\begin{pmatrix} I_{km}^{\mathrm{R}}(t) \\ I_{mk}^{\mathrm{R}}(t) \end{pmatrix} \ = \ \begin{pmatrix} I_{km}^{\mathrm{R}}(0) \\ I_{mk}^{\mathrm{R}}(0) \end{pmatrix} + \begin{pmatrix} \boldsymbol{\mu_{km}(t)} \\ \boldsymbol{\mu_{mk}(t)} \end{pmatrix} \qquad (7)$$

where $\boldsymbol{\mu_{km}(t)}$ and $\boldsymbol{\mu_{mk}(t)}$ are drawn from zero mean distributions with small standard deviation.

[a] We use boldface to indicate random variables.

As a functionally equivalent alternative to (7), the attacker could simply set

$$\begin{pmatrix} I_{km}^{\mathrm{R}}(t) \\ I_{mk}^{\mathrm{R}}(t) \end{pmatrix} \ = \ Y_{km} \begin{pmatrix} V_k^{\mathrm{R}}(t) \\ V_m^{\mathrm{R}}(t) \end{pmatrix} \qquad (8)$$

our analyses in the following apply to either form.

A second form of attack that has been considered is the *data replay* attack. Here, the attacker supplies a previously observed (or computed) pair of time series $V^{\mathrm{R}}(t)$ and $I^{\mathrm{R}}(t)$ for buses and lines within the set $\mathcal{A}$, which agrees at $t = 0$ with the values $V^{\mathrm{R}}(0)$ and $I^{\mathrm{R}}(0)$ computed as previously.

1) Discussion: In the abovementioned template, the attacker wishes to inject a semblance of randomness in the reported data while attempting to keep the attack undetected, ideally for a period measured in minutes.

We assume the attacker chooses distributions for the $\boldsymbol{\nu_k(t)}$, $\boldsymbol{\mu_{km}(t)}$, and $\boldsymbol{\mu_{mk}(t)}$ that have sufficiently small *standard deviation* relative to $V_k^{\mathrm{R}}(0)$, $I_{km}^{\mathrm{R}}(0)$, and $I_{mk}^{\mathrm{R}}(0)$, respectively. This choice is justified as follows. Recall that the attack is assumed to take place during a period of slow load changes. Thus, (true) voltages, currents, and power flows will all be slowly changing. Also recall that the initial attack computation enforces (exact) current–voltage and power-injection consistency in the initial attack. Thus, if the standard deviations in Template V.1 are as stated, both current–voltage and power-injection consistency will approximately hold following the initial attack, with high probability.

VI. Defense

As discussed above, prior work and our computations in Section VIII-A show that it is possible to compute high-fidelity attacks that disguise dangerous network conditions, even in large, complex transmission systems.

In this section, we describe a generic randomized defense strategy that can deployed when a complex attack is suspected but has not been verified. The application of real-time monitoring methods such as those in [29] or [30], or similar, could be used as a filter to suggest that an attack or unexplained event has taken place, thereby triggering the defensive mechanism. In the analyses, we will assume that the attack impacts a proper subset $\mathcal{A}$ of the system that is unknown to the control center, as was the case mentioned above; although the generic defense strategy applies under more general attacks as well.

Assumption VI.1: We will assume that if there is a topology change the network remains connected.

The strategy can be summarized by the following template.

Procedure VI.2: Random injection defense.

Iterate:

D1: Choose, for each $k \in \mathcal{G}$ a (random) value $\boldsymbol{\delta_k}$ such that $\sum_{g \in \mathcal{G}} \boldsymbol{\delta_k} = 0$. Command each generator $k \in \mathcal{G}$ to change its output to $P_k^g + \boldsymbol{\delta_k}$.

D2: Following the generation change in step **D1** identify inconsistencies in the observed sensor readings.

Here, an "inconsistency" is an incorrect condition satisfied by the reported data (such as a violation of current–voltage consistency or power-injection consistency), or stochastic behavior that is inconsistent with system-wide behavior understood by the control center.

We will describe several concrete versions of this idea as follows. See Procedures VI.4, VI.5, and VII.1.

Each iteration would last several seconds, and statistically significant inconsistencies identified by this scheme are flagged as potential evidence of an attack. In the following, we will describe several specific implementations of the random ingredient. The strategy in Procedure VI.2 is likely to succeed, in particular against the noisy data or data replay attacks, if the generation changes result in significant voltage changes across the system. Lemma 1 explains why a particular implementation of Procedure VI.2 attains this goal.

We note that there is an existing literature on using network resources so as to change power flow physics in order to detect structure or faults (see [32]–[36]). Indeed, even though the description of our random injection defense focuses on power injections, one could also consider other random probing strategies that change power flows, such as adjusting transformer settings, controlled line tripping, and the use of other technologies such as storage and solid-state devices.

There are several implementations of the generic strategy. Generally the defender wants to make the $|\boldsymbol{\delta_k}|$ large because to first-order changes in voltage angles are proportional to $\|\boldsymbol{\delta}\|_2$, and a large change in phase angles is likely to give rise to a significant current–voltage or power-injection inconsistencies in sensor readings in $\partial\mathcal{A}$, as discussed previously. This idea forms the basis for a simple, current-consistency-based version of Template VI.2 (see Procedure VI.4).

An attacker, who is aware that the random injection defense strategy is applied, may try to replace, e.g., the noisy data attack with a more careful manipulation of reported data. For example, the attacker could react to a significant change to voltages in $\partial\mathcal{A}$ by solving a nonlinear, the nonconvex system of inequalities designed to guarantee approximate current–voltage and power-injection consistency. In addition, any net load change within $\mathcal{A}$ implied by the manipulated data must be very small (or it could contradict observed frequencies). Finally the attacker would need to perform this computation very quickly, and repeatedly (because the defense will be applied repeatedly).

This online complex computation could in principle be bypassed by the attacker by considering changes to readings of voltages at buses in $\partial\mathcal{A}$ only; with the remaining voltages in $\mathcal{A}$ computed as in Template V.1. We will term this (hypothetical) attack as the *enhanced* noisy data attack. However, in Section VI-B, we will show that when the random injection defense causes large-enough voltage changes in $\partial\mathcal{A}$, the enhanced noisy-data attack fails, even accounting for realistic sensor error (See Lemma 7).

Our defensive strategies can be easily adjusted if sensors are not available throughout the system, by restricting the tests we perform to sensorized buses and lines. Of course, the fewer the sensors are the more limited the impact of the defense is. Indeed, some interesting work (using the standard, dc-equation state estimation) precisely seeks to perform system identification post attack when only limited sensor information is available [7]–[10].

A. Controlling Voltages Through Generation Changes

As discussed previously, a goal of the defense is to produce large voltage angle changes in buses in $\mathcal{A}$, with the intention of revealing inconsistencies in reported data on lines between $\partial\mathcal{A}$ and $\mathcal{A}^c$, defined as the complement of $\mathcal{A}$. The defender, of course, does not know the set $\mathcal{A}$, and thus, it is of interest to understand when the voltage at any given bus can be changed by appropriately choosing the injections δ.

In this section, we address the task of changing voltage angles through injections through Procedure VI.2. We will argue, using the dc power flow approximation (1), that a specific version of Procedure VI.2 does succeed in this task (see Lemmas 1 and 3). In Section VI-A1, we will present experiments under the ac power model that verify the dc-based results. And in Section VI-B, we further argue that the voltage changes are large enough to overcome sensor error.

We begin with a useful technical point summarized by (10b). Denote by $\hat{B}$, the bus susceptance matrix of the network, post attack. This matrix will be different from the original bus susceptance matrix, B, in case of a topology or susceptance attack; thus, the control center does not know $\hat{B}$. Recall that as stated previously, we are assuming that the network remains connected, post attack. Suppose that we consider an iteration of

Procedure VI.2. Define

$$\hat{\boldsymbol{P}}^{\boldsymbol{g}} \ = \ P^g + \boldsymbol{\delta}$$

this is the vector of generation injections as per Procedure VI.2.

Consider the two systems

$$\hat{B}\boldsymbol{\theta} = P^g - P^d \tag{9a}$$

$$\hat{B}\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{P}}^{\boldsymbol{g}} - P^d \tag{9b}$$

which give the phase angles before (resp., after) an application of the defense. The matrix $\hat{B}$, under the connectedness assumption, has rank equal to the number of rows, minus 1. Hence, we can pick an arbitrary *reference bus* t, i.e., we can set (without loss of generality) $\theta_t = 0$ (resp., $\hat{\boldsymbol{\theta}}_{\boldsymbol{t}} = 0$), in which case the solution to either system (9a), (9b) is unique, and given by

$$\theta = \breve{B}_t(P^g - P^d) \tag{10a}$$

$$\hat{\boldsymbol{\theta}} = \breve{B}_t(\hat{\boldsymbol{P}}^{\boldsymbol{g}} - P^d) \ = \ \theta + \breve{B}_t\boldsymbol{\delta} \tag{10b}$$

where $\breve{B}_t$ is an appropriate pseudo-inverse of $\hat{B}$. We will rely on (10b).

Important observation: Let i and j be two arbitrary buses. Let θ and $\hat{\boldsymbol{\theta}}$ be as in (10). Then, the two quantities $\theta_i - \theta_j$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{j}}$ are *invariant under the choice of the reference bus* t.

The defensive strategy that we develop, as a specific implementation of the random injection Procedure VI.2, assumes that there is a known set $\mathcal{T}$ of *generator* buses that are known to be "trusted," that is to say, we can assume that data from buses in $\mathcal{T}$ is known to be unmodified. This concept is not new; see [6], [22], [37], [38] for related discussions. Without such an assumption the entire suite of signals received by the control centers could be falsified and it is questionable whether any meaningful attack reconstruction can be performed.

Assumption VI.3: $|\mathcal{T}| \geq 2$.

The following template describes the strategy:

Procedure VI.4: Pairs-driven version of Procedure VI.2.

At each execution of step **D1**, choose an ordered pair of buses (s, t) in $\mathcal{T}$, selected with equal probability from among all such pairs of buses. Select random $\boldsymbol{\Gamma} > 0$, and use the injection perturbations

$$\boldsymbol{\delta_s} = \boldsymbol{\Gamma}, \ \boldsymbol{\delta_t} = -\boldsymbol{\Gamma}, \text{ and } \boldsymbol{\delta_k} = 0 \quad \forall k \neq s, t. \tag{11}$$

Remark: In the abovementioned procedure, any stochastic distribution for $\boldsymbol{\Gamma}$ can be used. In Section VII, we will return to this point.

Lemmas 1 and 2, given next, consider an iteration of Procedure VI.4, and give structural properties to be used in the sequel.

Lemma 1: Suppose $\hat{B}\theta = P^g - P^d$, and $\hat{B}\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{P}}^{\boldsymbol{g}} - P^d$. Let $k \neq t$ be a bus such that the postattack network contains a path between s and k that does not include t. Then

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{t}} \ > \ \theta_k - \theta_t. \tag{12}$$

Proof: As discussed previously [see (9)], we may assume $\hat{\boldsymbol{\theta}}_{\boldsymbol{t}} = \theta_t = 0$. Under this assumption (12) reads

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \theta_k > 0. \tag{13}$$

Let M be the set of buses $p \neq t$ such that

1) the network contains a path from s to p that avoids t; and
2) subject to (1), $\hat{\boldsymbol{\theta}}_{\boldsymbol{p}} - \theta_p$ is *minimum*.

Aiming for a contradiction, we will assume that

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{p}} - \theta_p \ \leq \ 0 \quad \text{for } p \in M. \tag{14}$$

Showing that (14) is false yields (13). For any line km define the flow value $f_{km} = (\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{m}} - \theta_k + \theta_m)/x_{km}$. Since $\hat{B}(\hat{\boldsymbol{\theta}} - \theta) = \hat{\boldsymbol{P}}^{\boldsymbol{g}} - P^g$, the flow vector f corresponds (under the dc power flow model) to a power flow with $\boldsymbol{\Gamma}$ units of generation at s, $\boldsymbol{\Gamma}$ units of load at t, and zero generation and load elsewhere. Note that for any line km, $f_{km} > 0$ iff

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \theta_k > \hat{\boldsymbol{\theta}}_{\boldsymbol{m}} - \theta_m. \tag{15}$$

This observation implies

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{s}} - \theta_s > 0. \tag{16}$$

[To obtain this fact, decompose the flow vector f into a set of path flows from s to t and telescope (15) along any such path.] Pick any $p \in M$ and let P be a path from s to p that avoids t. Say $P = v_0, v_1, \ldots, v_i$ where $v_0 = s$ and $v_i = p$, and let h be smallest such that $v_h \in M$. By (16), $s \notin M$, i.e., $h > 0$. Then, by definition of h, $\hat{\boldsymbol{\theta}}_{\boldsymbol{v_{h-1}}} - \theta_{v_{h-1}} > \hat{\boldsymbol{\theta}}_{\boldsymbol{v_h}} - \theta_{v_h}$, i.e., $f_{v_{h-1}, v_h} > 0$. But by assumption $v_h \neq t$. So there exists some line v_h, m such that $f_{v_h, m} > 0$. Therefore, using the assumption $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \theta_k \leq 0$ for all $k \in M$, $v_h \in M$, and (15)

$$0 \geq \hat{\boldsymbol{\theta}}_{\boldsymbol{v_h}} - \theta_{v_h} > \hat{\boldsymbol{\theta}}_{\boldsymbol{m}} - \theta_m. \tag{17}$$

So $m \neq t$, and as a result by construction there is a path from s to m that avoids t. But then (17) contradicts the fact that $v_h \in M$. ■

Remark: if $\boldsymbol{\Gamma}$ is chosen negative in (11), then instead of (12), we obtain $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{t}} \ < \ \theta_k - \theta_t$ through essentially the same proof.

For future reference, we state the following analogue of Lemma 1, with similar proof (omitted).

Lemma 2: Suppose θ and $\hat{\boldsymbol{\theta}}$ be as in Lemma 1. Let $k \neq t$ be a bus such that in the postattack network every path between s and k must include t. Then

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{t}} \ = \ \theta_k - \theta_t. \tag{18}$$

Lemmas 3 and 4, given next, will be used to construct a specific detection criterion (see Procedure VI.5) given in the following.

Lemma 3: Let $k \notin \mathcal{T}$ be any bus. Let u, v be any two buses in $\mathcal{T}$. Then, in the postattack network, at least one of the following conditions apply:

i) some path between u and k does not include v;
ii) some path between v and k does not include u.

Proof: Without loss of generality, in the post attack network the distance between u and k is no larger than the distance between v and k. Then, (i) holds. ■

To analyze the pairs-driven defense during a particular iteration of Procedure VI.4, involving pair (s, t), we will express phase angles using t as the reference bus; see (10b). As a result of Lemmas 1 and 2 we have the following.

Lemma 4: Let $k \notin \mathcal{T}$. Consider any iteration of Procedure VI.4. Then (a)

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} = \theta_k + \beta_k \boldsymbol{\Gamma} \tag{19}$$

for some $\beta_k \geq 0$. (b) $\beta_k > 0$ with probability at least $1/2$.

Proof: (a) Equation (19) follows from (10b) and $\beta_k \geq 0$ follows from Lemmas 1 and 2. (b) Under Procedure VI.4, the probability that an ordered pair (s, t) is chosen is equal to the probability that (t, s) is chosen. As a result, by Lemma 3, with probability at least $1/2$ an ordered pair (s, t) is chosen so that there is a path between s and k that does not include t. In such a case by Lemma 1, we will have $\beta_k > 0$. ■

Corollary 5: Let $k \notin \mathcal{T}$ and consider L iterations of Procedure VI.4. The probability that $\beta_k = 0$ for every iteration is at most $1/2^L$.

Lemma 4 suggests the following detection paradigm.

Procedure VI.5: Pairs-driven detection criterion.

As Procedure VI.4 iterates, for each bus k, estimate the correlation between $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ and $\boldsymbol{\Gamma}$. The defensive procedure terminates when all these estimates are stable. At that point, any bus k whose correlation coefficient is nonpositive is flagged as suspicious.

Lemma 6: Suppose the pairs-driven defense is operated for $L > 0$ iterations. As $L \to +\infty$, the probability the pairs-driven defense will defeat the noisy data and data replay attacks converges to 1. More precisely, with probability converging to 1 each bus whose data are modified will be flagged, and any bus that is not attacked will not be flagged.

Proof: Let $k \notin \mathcal{T}$. Then, by Corollary 5 with probability at least $1 - 1/2^L$, the control center expects to see at least one iteration during which $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}} = \theta_k + \beta_k \boldsymbol{\Gamma}$ for some value $\beta_k > 0$ (unknown to the control center). This fact yields the desired result since in either attack case, if $k \in \mathcal{A}$, the correlation between $\hat{\boldsymbol{\theta}}_{\boldsymbol{k}}$ and $\boldsymbol{\Gamma}$ will be zero. ■

Lemmas 1 through 6 assume the dc power flows model, which is only a first-order approximation to the ac model we consider here. In the following section, we perform numerical experiments, under the ac power flows model, of the random injection defense VI.2. A separate issue concerns ambient *noise*; we need voltage changes to overcome noise levels in measurements. This issue is taken up in Section VI-B.

1) Numerical Experiments Using AC Power Flows: The abovementioned discussion concerns dc power flows. In order to investigate how voltages change under injection changes, under ac power flows, we perform a experiments using examples from the Matpower library [39]. For each system we perform ten experiments. In each experiment, we compute an ac power flow, which is constrained to satisfy the given voltage bounds at all generator buses, but not at load buses, as well as power injection constraints and generator limits, while allowing large injection changes in a random subset of generators. For a nongenerator bus k, let V_k^b be its voltage in the base case (i.e., the Matpower case), and let $V_{i,k}$ be its voltage in experiment $i = 1, \ldots, 10$. Finally, define

$$\text{score}(k) \doteq \max_{1 \leq i \leq 10} \frac{|V_{i,k} - V_k^b|}{|V_k^b|}.$$

TABLE I
AC VOLTAGE CHANGES

Case	Min Score	Average Score
case118	11.61%	32.77%
case1354pegase	7.62%	51.00%
case2746wp	5.00%	10.09%

In Table I, "Min Score" is the minimum score across all nongenerator buses. Thus, the table provides experimental verification for substantial ac voltage changes under random generator injections.

B. Overcoming Sensor Error, and the Current–Voltage Defense

If sensor misestimation (i.e., *error*) is present, a strategy based on Procedure VI.2 may fail to detect data inconsistencies if the random power injections cause voltage changes that are too small as compared to the error. In order to derive a version of Procedure VI.2 that deals with this issue, we next describe a particular implementation of step **D2**, which relies on the current–voltage consistency condition (4), which takes into account the possibility of sensor error. Whereas abovementioned a phasor (voltage or current quantity) ϕ had a true physical value ϕ^{T} and a reported value ϕ^{R} (which is the value received by the control center), now we will have the *sensed* value ϕ^{S}, which is the value actually produced by the sensor.

Due to sensor error, sensed and true data may differ. For a phasor ϕ define $\mathrm{err}(\phi) \doteq \phi^{\mathrm{S}} - \phi^{\mathrm{T}}$. In the PMU setting, the total vector error (TVE) criterion [40], [41] guarantees that

$$|\mathrm{err}(\phi)| < \tau |\phi^{\mathrm{T}}| \tag{20}$$

where $0 < \tau < 1$ is a tolerance. Standards enforce $\tau = 1\%$, though experimental testing of PMUs shows far smaller errors [42]. From (20), we obtain

$$(1 - \tau)|\phi^{\mathrm{T}}| < |\phi^{\mathrm{S}}| < (1 + \tau)|\phi^{\mathrm{T}}| \tag{21a}$$

$$|\mathrm{err}(\phi)| < \tau(1 - \tau)^{-1}|\phi^{\mathrm{S}}|. \tag{21b}$$

We will describe two sensor-error-aware voltage–current consistency criteria. An important point is that the current–voltage consistency condition (4), combined with estimations of possible sensor error, yields a nonlinear relationship, and an appropriate reformulation of this relationship can render useful benefits. For a line km write

$$Y_{km} = \begin{pmatrix} Y_{km}^{(1)} & Y_{km}^{(2)} \\ Y_{km}^{(3)} & Y_{km}^{(4)} \end{pmatrix}.$$

Criterion 1: We have that $I^{\mathrm{T}}_{mk} = Y^{(3)}_{km} V^{\mathrm{T}}_{k} + Y^{(4)}_{km} V^{\mathrm{T}}_{m}$. Write $Z^{(3)}_{km} \doteq [Y^{(3)}_{km}]^{-1}$. Hence

$$\begin{aligned} &V^{\mathrm{S}}_{k} - Z^{(3)}_{km}(I^{\mathrm{S}}_{mk} - Y^{(4)}_{km} V^{\mathrm{S}}_{m}) \\ &\quad = \mathrm{err}(V_k) \; - \; Z^{(3)}_{km}(\mathrm{err}(I_{mk}) - Y^{(4)}_{km}\mathrm{err}(V_m)). \end{aligned} \tag{22}$$

Using the triangle inequality, we obtain

$$\begin{aligned} &|V^{\mathrm{S}}_{k} - Z^{(3)}_{km}(I^{\mathrm{S}}_{mk} - Y^{(4)}_{km} V^{\mathrm{S}}_{m})| \\ &\quad < |\mathrm{err}(V_k)| + |Z^{(3)}_{km}|(\,|\mathrm{err}(I^{\mathrm{S}}_{mk})| + |Y^{(4)}_{km}||\mathrm{err}(V_m)|\,) \end{aligned} \tag{23}$$

which yields, using (20) on the first term and (21b) on the last two terms of (23)

$$\begin{aligned} &|V^{\mathrm{S}}_{k} - Z^{(3)}_{km}(I^{\mathrm{S}}_{mk} - Y^{(4)}_{km} V^{\mathrm{S}}_{m})| \\ &\quad \leq \tau |V^{\mathrm{T}}_{k}| + \frac{\tau |Z^{(3)}_{km}|}{1-\tau}(|I^{\mathrm{S}}_{mk}| + |Y^{(4)}_{km}|\,|V^{\mathrm{S}}_{m}|). \end{aligned}$$

This last expression equals (using the current–voltage relationship)

$$\tau |Z^{(3)}_{km}(I^{\mathrm{T}}_{mk} - Y^{(4)}_{km} V^{\mathrm{T}}_{m})| + \frac{\tau |Z^{(3)}_{km}|}{1-\tau}(|I^{\mathrm{S}}_{mk}| + |Y^{(4)}_{km}|\,|V^{\mathrm{S}}_{m}|)$$

which, using the left inequality in (21a) and the triangle inequality, is strictly less than

$$\frac{2\tau |Z^{(3)}_{km}|}{1-\tau}(|I^{\mathrm{S}}_{mk}| + |Y^{(4)}_{km}|\,|V^{\mathrm{S}}_{m}|).$$

In summary

$$|V^{\mathrm{S}}_{k} - Z^{(3)}_{km}(I^{\mathrm{S}}_{mk} - Y^{(4)}_{km} V^{\mathrm{S}}_{m})| < \frac{2\tau |Z^{(3)}_{km}|}{1-\tau}(|I^{\mathrm{S}}_{mk}| + |Y^{(4)}_{km}||V^{\mathrm{S}}_{m}|). \tag{24}$$

Under Criterion 1, if, statistically, the reported phasors V^{R}_{k}, V^{R}_{m}, I^{R}_{mk} fail to satisfy (24),[3] line km is flagged as suspicious. A similar analysis concerns V^{R}_{k}, V^{R}_{m}, and I^{R}_{km}.

Remark: By construction, if $k, m \notin \mathcal{A}$, then line km will not be flagged.

Criterion 2: Proceeding as mentioned above, we have

$$\begin{aligned} &|I^{\mathrm{S}}_{km} - Y^{(1)}_{km} V^{\mathrm{S}}_{k} - Y^{(2)}_{km} V^{\mathrm{S}}_{m}| \\ &\quad < \frac{\tau}{1-\tau}(|I^{\mathrm{S}}_{km}| + |Y^{(1)}_{km}||V^{\mathrm{S}}_{k}| + |Y^{(2)}_{km}||V^{\mathrm{S}}_{m}|). \end{aligned} \tag{25}$$

(and similarly with I_{mk}).

Note: this criterion can be sharpened when line km is a pure impedance line (no transformer). If the reported phasors do not satisfy (25), then the line is flagged.

1) Discussion: Note that a line not attacked will not be flagged, as per the TVE condition. Additional criteria can be developed to handle power-injection consistency. To analyze the effectiveness of these criteria, we turn to the *enhanced* noisy data attack discussed in Section V-1. To remind the reader, in this type of attack, the voltage readings in $\partial\mathcal{A}$ can be arbitrarily adjusted. While this action may create inconsistencies on lines with just one end in $\partial\mathcal{A}$, the attacker may be able to "hide" such inconsistencies if they are small enough relative to sensor error.

We next show that Criterion 1 alone can suffice to defeat the enhanced noisy data attack (i.e., uncover inconsistencies) when voltage angles are sufficiently changed under our random injection defense.

To understand these points, consider a bus $k \in \partial\mathcal{A}$ such that there is a line km with $m \notin \mathcal{A}$ and also a line ka where $a \in \mathcal{A} - \partial\mathcal{A}$. We study an iteration of the random injection defense that (to simplify notation) we assume begins at time $t = 0$. Consider line ka first. To avoid having line ak flagged, the attacker t will need to manufacture time series $V^{\mathrm{R}}_{k}(t)$, $V^{\mathrm{R}}_{a}(t)$, and $I^{\mathrm{R}}_{ak}(t)$ that (statistically) satisfy (24). But under the noisy data attack, on average $V^{\mathrm{R}}_{a}(t) = V^{\mathrm{R}}_{a}(0)$ and $I^{\mathrm{R}}_{ak}(t) = I^{\mathrm{R}}_{ak}(0)$. Hence, to defeat Criterion 1, the attacker needs (on average) that

$$\begin{aligned} &\frac{2\tau |Z^{(3)}_{ka}|}{1-\tau}(|I^{\mathrm{R}}_{ak}(0)| + |Y^{(4)}_{ka}|\,|V^{\mathrm{R}}_{a}(0)|) \\ &\; > |V^{\mathrm{R}}_{k}(t) - Z^{(3)}_{ka}(I^{\mathrm{R}}_{ak}(0) - Y^{(4)}_{ka} V^{\mathrm{R}}_{a}(0))| = |V^{\mathrm{R}}_{k}(t) - V^{\mathrm{R}}_{k}(0)|. \end{aligned} \tag{26}$$

Now consider line km. Since $m \notin \mathcal{A}$, $V^{\mathrm{R}}_{m}(t) = V^{\mathrm{S}}_{m}(t)$, and $I^{\mathrm{R}}_{mk}(t) = I^{\mathrm{S}}_{mk}(t)$. Also, denote the following points.

1) $V^{\mathrm{T}}_{k}(*) =$ the true voltage at k at the start of the current iteration of the random injection defense, i.e., the voltage resulting from the injection changes in step **D1**. Then, assuming unbiased sensor errors and zero-mean ambient noise, $V^{\mathrm{T}}_{m}(*)$ will equal the expectation of $V^{\mathrm{T}}_{m}(t)$ during the iteration.
2) Likewise define the current $I^{\mathrm{T}}_{mk}(*)$.

Hence, to defeat Criterion 1, the attacker needs (on average) that

$$\begin{aligned} &\frac{2\tau |Z^{(3)}_{km}|}{1-\tau}(|I^{\mathrm{T}}_{mk}(*)| + |Y^{(4)}_{km}|\,|V^{\mathrm{T}}_{m}(*)|) \\ &\quad > |V^{\mathrm{R}}_{k}(t) - Z^{(3)}_{km}(I^{\mathrm{T}}_{mk}(*) - Y^{(4)}_{km} V^{\mathrm{T}}_{m}(*))| \\ &\quad = |V^{\mathrm{R}}_{k}(t) - V^{\mathrm{T}}_{k}(*)|. \end{aligned} \tag{27}$$

As a result of these observations we have the following.

Lemma 7: Consider buses k, a, m as described previously. Suppose that

$$\begin{aligned} |V^{\mathrm{T}}_{k}(*) - V^{\mathrm{R}}_{k}(0)| > &\frac{2\tau |Z^{(3)}_{ka}|}{1-\tau}(|I^{\mathrm{R}}_{ak}(0)| + |Y^{(4)}_{ka}|\,|V^{\mathrm{R}}_{a}(0)|) \\ &+ \frac{2\tau |Z^{(3)}_{mk}|}{1-\tau}(|I^{\mathrm{T}}_{mk}(*)| + |Y^{(4)}_{km}|\,|V^{\mathrm{T}}_{m}(*)|). \end{aligned} \tag{28}$$

Then, it is impossible for the enhanced noisy data attacker to statistically satisfy Criterion 1 on both lines ka and km.

Proof: As argued previously, the attacker needs both (26) and (27) to hold. Since

$$|V^{\mathrm{T}}_{k}(*) - V^{\mathrm{R}}_{k}(0)| \leq |V^{\mathrm{R}}_{k}(t) - V^{\mathrm{R}}_{k}(0)| + |V^{\mathrm{R}}_{k}(t) - V^{\mathrm{T}}_{k}(*)|$$

summing (26) and (27), we obtain a contradiction to (28). ■

Comment: This lemma highlights how large changes in voltages caused by the random injection defense challenge the attacker.

[3] More precisely: if the expected difference between the right-hand and left-hand side of (24) is nonpositive.

TABLE II
CURRENT–VOLTAGE DEFENSE

	Experiment 1	Experiment 2
ϵ	0.01	0.05
$\sum_{k \in \mathcal{G}} \delta_k^+$	289.01	964.77
$\sum_{k \in \mathcal{G}} \delta_k^-$	174.47	256.04
Line $(k = 1139, a = 1137)$		
$\lvert V_a^{\mathrm{R}}(0)\rvert \angle \theta_a^{\mathrm{R}}(0)$	$1.0919\angle - 6.993°$	$1.0919\angle - 6.993°$
$I_{ak}^{\mathrm{R}}(0)$	$-0.0275 + 0.0281j$	$-0.0275 + 0.0281j$
Line $(k = 1139, m = 1110)$		
$\lvert V_m^{\mathrm{T}}(*)\rvert \angle \theta_m^{\mathrm{T}}(*)$	$1.0309\angle - 7.822°$	$1.0391\angle - 7.848°$
$I_{mk}^{\mathrm{T}}(*)$	$0.0905 - 0.4976j$	$0.1289 - 0.4901j$
Voltages at $k = 1139$		
$\lvert V_k^{\mathrm{R}}(0)\rvert \angle \theta_k^{\mathrm{R}}(0)$	$1.0919\angle - 6.991°$	$1.0919\angle - 6.991°$
$\lvert V_k^{\mathrm{T}}(*)\rvert \angle \theta_k^{\mathrm{T}}(*)$	$1.0104\angle - 7.822°$	$1.0187\angle - 7.936°$
Lemma 7 applied to bus $k = 1139$		
Ratio	1.913	1.732

2) Experiment: Next, we describe a set of experiments involving the current–voltage defense applied to the attack given in the Appendix. The current defense was implemented as follows.

1) The set of responding generators, $\mathcal{R}$, was of cardinality 200. For $k \in \mathcal{R}$ $|\delta_k|$ can be arbitrarily large. We chose $\delta_k > 0$ with probability $1/2$.
2) For any generator bus $k \notin \mathcal{R}$, $|\delta_k| \leq \epsilon P_k^g$. We used values $\epsilon = 0.01, 0.05$.
3) No generator may exceed its limits (voltage or generation), but subject to all these conditions we maximize $\sum_{k \in \mathcal{G}} |\delta_k|$.

In Table II, we perform the abovementioned analysis on the lines $(k = 1139, a = 1137)$ and $(k = 1139, m = 1110)$ with $\tau = 0.01$. "Ratio" is the ratio of the left-hand side to the right-hand side of expression (28). We see that the condition for Lemma 7 is amply satisfied. A similar analysis pertains to line (1141, 1361), the other line connecting $\mathcal{A}$ to its complement.

VII. Covariance Defense

In this section, we describe an elaboration of the pairs-driven defense VI.4; the elaboration is motivated by the fact that real-world PMU data streams exhibit nongeneric stochastic structure in (for example) voltage angles [26], [43], [44]. In particular, covariance matrices across several time scales have very low rank (typically smaller than 10). Our defense will defeat both the noisy-data and data-replay attacks, under appropriate assumptions.

As mentioned above, we assume that the buses in a certain set $\mathcal{T}$ are *trusted* and that $|\mathcal{T}| \geq 2$. The emphasis of the methods in this section is that we aim to modify the *covariance* matrix of phase angles, whereas the random injection defense in Procedures VI.2 or VI.4 change the *average* voltage angle values. Such a change should prove more difficult for the attacker to correctly counteract since such a correction involves an estimation that requires time, during which the attacker will be producing incorrect data.

We additionally assume that the attacker's data stochastics are *stationary* (i.e., the parameters of the stochastic process do not change as a function of time). This implies in particular that the attacker does not react to the covariance defense by changing the stochastics of reported (falsified) data at attacked buses. In the following, we will discuss, however, why reacting to the defense would prove very difficult. We also assume that ambient conditions are also stationary.

To describe the defense we need some definitions. For an ordered pair of buses $(s, t) \in \mathcal{T}$, define the vectors $u^{s,t}$ and $v^{s,t}$ by

$$u_s^{s,t} = 1, \; u_t^{s,t} = -1, \text{ and } u_k^{s,t} = 0 \quad \forall\, k \neq s, t \tag{29a}$$

$$v^{s,t} \doteq \breve{B}_t u^{s,t}. \tag{29b}$$

Formally, the covariance defense works as follows. Let t_1, t_2 be two fixed trusted buses, and let $\mathcal{P}$ be a real-valued, zero-mean, positive variance probability distribution. The defense has the following two phases.

1) During an initial phase, post the suspected attack, for $i = 1, 2$, we compute the matrix.

$\sigma^2_{\boldsymbol{\theta}^{\mathrm{R}}, i}$ = covariance matrix of observed phase angles, expressed with respect to reference bus t_i.

2) After the initial phase, we perform iterations as in Procedure VI.4, with two modifications. First, we select an ordered pair of buses of the form (s, t) where $t = t_1$ or $t = t_2$ and $s \in \mathcal{T} - t$. This selection is made with equal probability from among all pairs that have the form we have just described. Second, the quantity $\boldsymbol{\Gamma}$ is drawn from the distribution $\mathcal{P}$, independently from observed (including ambient) stochastics. We then apply step **D1** of Procedure VI.4 using this triple $(s, t, \boldsymbol{\Gamma})$. Note that possibly $\boldsymbol{\Gamma} < 0$ and in this respect we are modifying Procedure VI.4. For $i = 1, 2$, an iteration where $t = t_i$ will be said to be of type-i.

This description concludes our modification to Procedure VI.4; next we describe the proposed detection criterion.

Throughout the second phase, for $i = 1, 2$, at each iteration of type-i we compute the matrix.

1) $\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{R}}, i}$ = sample covariance matrix of observed phase angles with respect to reference bus t_i.

The computation of $\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{R}}, i}$ uses the time series consisting of all prior data involving iterations of type-i since the start of the defense. The defense concludes when the estimates for these two matrices become stable.

Procedure VII.1: Covariance-driven detection criterion.

At termination of the defense, we flag a bus k as *suspicious* if, for either $i = 1$ or $i = 2$, the difference between the (k, k) entry of $\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{R}}, i}$ and the corresponding entry of $\sigma^2_{\boldsymbol{\theta}^{\mathrm{R}}, i}$ is smaller than λ, defined by

$$\lambda \doteq \frac{\sigma^2_{\boldsymbol{\Gamma}}}{|\mathcal{T}| - 1}\omega, \text{ where} \tag{30a}$$

$$\omega \doteq \min_{s,t,j} \{(v_j^{s,t})^2 \; : \; v_j^{s,t} \neq 0\}. \tag{30b}$$

This concludes the description of the defense, with analysis given in Lemmas 8, 10, and 6.

In preparation for those results, for $i = 1, 2$, we consider the random variables $\hat{\boldsymbol{\theta}}^{\mathbf{T},i}$ and $\boldsymbol{\theta}^{\mathbf{T},i}$, which are defined during iterations of type-i:

(a.1) $\hat{\boldsymbol{\theta}}^{\mathbf{T},i}$ is the vector of true voltage phase angles, using t_i as the reference bus.

(a.2) $\boldsymbol{\theta}^{\mathbf{T},i}$ describes the true vector of voltage phase angles, had the power injections in the defense *not* been applied at that point of time. It is also given using t_i as the reference bus.

Note: neither random variable can actually be observed by the control center; the above mentioned are mathematical definitions. Lemma 8 given next (similar to Lemma 4) concerns the relationship between these two random variables.

Lemma 8: For $i = 1, 2$, consider an iteration of type-i. Let (s, t_i) be the pair being used. Then, $\hat{\boldsymbol{\theta}}^{\mathbf{T},i} = \boldsymbol{\theta}^{\mathbf{T},i} + \boldsymbol{\Gamma} v^{s,t_i}$.

Proof: Given that the pair (s, t_i) is being used, the injections in the random defense, per (11) are given by $\boldsymbol{\delta} = \boldsymbol{\Gamma} u^{s,t_i}$. The result follows from (10b). ■

The following result presents a key feature of the covariance of phase angles. Recall that $\boldsymbol{\Gamma}$ is drawn independent of all other stochastics.

Lemma 9: For $i = 1, 2$,

$$\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{T},i}} = \sigma^2_{\boldsymbol{\theta}^{\mathrm{T},i}} + \frac{\sigma^2_{\boldsymbol{\Gamma}}}{|\mathcal{T}| - 1} \sum_{s \in \mathcal{T} - t_i} v^{s,t_i} (v^{s,t_i})^{\top}. \tag{31}$$

Proof: We proceed by conditioning on iterations of type-i where a particular pair (s, t_i) is selected by the defense. Subject to this conditioning, by Lemma 8, the covariance of $\hat{\boldsymbol{\theta}}^{\mathbf{T},i}$ equals

$$\sigma^2_{\boldsymbol{\theta}^{\mathrm{T},i}} + \sigma^2_{\boldsymbol{\Gamma}} v^i (v^i)^{\top} + \mathrm{covar}(\boldsymbol{\theta}^{\mathbf{T},i}, \boldsymbol{\Gamma} v^i).$$

The last term in this expression is zero, by the independence assumption on $\boldsymbol{\Gamma}$. The result follows since each pair is chosen with probability $(|\mathcal{T}| - 1)^{-1}$. ■

Lemma 10: Let k be any bus. Then, for at least one of $i = 1$ or 2, the (k, k) entry of $\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{T},i}}$ is at least as large as the corresponding entry of $\sigma^2_{\boldsymbol{\theta}^{\mathrm{T},i}}$, plus λ (defined as in (30a)).

Proof: Without loss of generality, there is a path between k and t_1 that avoids t_2. Note that the pair (s, t_2) with $s = t_1$ is one of the pairs available for the defense. By Lemma 1, we have that $v_k^{t_1,t_2} > 0$ and so $v_k^{t_1,t_2} \geq \omega^{1/2}$. Considering equation (31) for $i = 2$, we see that one of the terms in the sum corresponds to $s = t_1$. As just argued, the (k, k) entry of this term is at least ω. The (k, k) entries in the remaining terms of the sum are nonnegative (since each term is a positive-semidefinite matrix). Thus, the result follows. ■

Lemma 11: A bus is attacked if and only if it is flagged as suspicious by the covariance defense.

Proof: Consider first a bus k that is not attacked. For such a bus, by definition, $\hat{\boldsymbol{\theta}}^{\mathbf{T},i}{}_k = \hat{\boldsymbol{\theta}}^{\mathbf{R},i}{}_k$, for both $i = 1, 2$. Thus, by Lemma 10, bus k is not flagged as suspicious. On the other hand, suppose k is attacked. Then, under either the noisy-data or data-replay attacks the (k, k) entry of $\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{R},i}}$ will be equal to the corresponding entry of $\sigma^2_{\boldsymbol{\theta}^{\mathrm{R},i}}$, by the stationarity assumption (the attacker does not change stochastics when the defense is implemented). Hence, bus k is flagged. ■

Remarks: Given a pair (s, t_i) used in the defense, by Lemma 1 any entry v_k^{s,t_i} is positive if there is a path from bus k to s that avoids t_i. Let A denote the set of such buses. By Lemma 2, for $k \notin A$, $v_k^{s,t_i} = 0$.

Thus, in the term $v^{s,t_i}(v^{s,t_i})^{\top}$ in (31), the entire submatrix with rows and columns in A is positive, and the remaining entries in $v^{s,t_i}(v^{s,t_i})^{\top}$ are zero. By adjusting the proof of Lemma 10, we conclude that for k and m in A, the entry (k, m) of $\sigma^2_{\hat{\boldsymbol{\theta}}^{\mathrm{T},i}}$ is at least as large as the corresponding entry of $\sigma^2_{\boldsymbol{\theta}^{\mathrm{T},i}}$, plus λ. Thus, a submatrix of the covariance matrix will change via the defense (and not just the diagonal entries). If the network is guaranteed to be 2-connected [45] one can prove that the entire covariance matrix must change.

The covariance defense has an additional important feature, that the right-hand side of (31) has rank-$|\mathcal{T}| - 1$ whereas we expect the left-hand side of (31) to have low rank.

Lemma 12: For each $i = 1, 2$, the vectors $v^{s,t_i} = \breve{B} u^i$ as in (29b) are linearly independent. Hence, the second term in (31) has rank at least $|\mathcal{T}| - 1$.

Proof: The $|\mathcal{T}| - 1$ vectors u^{s,t_i} arising from all pairs (s, t_i) under consideration are linearly independent, by construction in (29a). Hence, the corresponding vectors v^{s,t_i} are also linearly independent. ■

Lemma 12 highlights the challenges faced by the attacker, even if the attacker is aware that the covariance defense is being deployed: the attacker will have to alter reported data in a way consistent with a nontrivial structural change in the covariance matrix. The attacker could "learn" the changes, by observing sensor output, but doing so would require time, during which the attacker is still expected to produce data readings, producing an error trail.

Appendix
Computation of Initial Attack

In this section, we describe a procedure that computes ac-undetectable attacks as per the discussion in Section IV; the attacks can involve line disconnection and load modification (which is actually computed). Other actions (impedance changes, transformer changes, etc) could also be incorporated.

As input to the computation, we have a set $\mathcal{A} \subset \mathcal{N} \backslash \mathcal{G}$ of buses (the target zone), a set of lines $\mathcal{L}$ to be disconnected, all with both ends in $\mathcal{A}$, and a line $uv \notin \mathcal{L}$ with both ends in $\mathcal{A}$. Write $\mathcal{A}^C = \mathcal{N} \setminus \mathcal{A}$. Let $(\hat{S}_k^g = \hat{P}_k^g + j\hat{Q}_k^g)_{k \in \mathcal{N}}$ and $(\hat{S}_k^d = \hat{P}_k^d + j\hat{Q}_k^d)_{k \in \mathcal{N}}$ be, respectively, the complex power generation and loads at the time of the attack. We assume that the attacker observes all these quantities. An explanation of the variables and constraints in the formulation given in the following:

$$\text{Max } (p_{uv}^{\mathrm{T}})^2 + (q_{uv}^{\mathrm{T}})^2 \tag{32a}$$

s.t.

$$\forall k \in \mathcal{A}^C \cup \partial\mathcal{A},\ |V_k^{\mathrm{T}}| = |V_k^{\mathrm{R}}|,\ \theta_k^{\mathrm{T}} = \theta_k^{\mathrm{R}} \tag{32b}$$

$$\forall k \in \mathcal{A},\ -(P_k^{d,\mathrm{R}} + jQ_k^{d,\mathrm{R}}) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{R}} + jq_{km}^{\mathrm{R}}) \tag{32c}$$

$$-(P_k^{d,\mathrm{T}} + jQ_k^{d,\mathrm{T}}) = \sum_{km \in \delta(k) \backslash \mathcal{L}} (p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}}) \tag{32d}$$

$$P_k^{d,\mathrm{R}} \geq 0,\ P_k^{d,\mathrm{T}} \geq 0 \tag{32e}$$

$\forall k \in \mathcal{A}^C \backslash \mathcal{R}$

$$\hat{P}_k^g - \hat{P}_k^d + j(\hat{Q}_k^g - \hat{Q}_k^d) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}}) \tag{32f}$$

$$\forall k \in \mathcal{R}: \qquad P_k^g - \hat{P}_k^g = \alpha_k \Delta \tag{32g}$$

$$P_k^g - \hat{P}_k^d + j(Q_k^g - \hat{Q}_k^d) = \sum_{km \in \delta(k)} (p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}}) \tag{32h}$$

$\forall k \in \mathcal{G}$

$$P_k^{g,\min} \leq P_k^g \leq P_k^{g,\max},\ Q_k^{g,\min} \leq Q_k^g \leq Q_k^{g,\max} \tag{32i}$$

$$\forall k \in \mathcal{N}: \ V_k^{\min} \leq |V_k^{\mathrm{T}}|, |V_k^{\mathrm{R}}| \leq V_k^{\max} \tag{32j}$$

$\forall$ line km :

$$|\theta_k^{\mathrm{R}} - \theta_m^{\mathrm{R}}| \leq \theta_{km}^{\max};\ |\theta_k^{\mathrm{T}} - \theta_m^{\mathrm{T}}| \leq \theta_{km}^{\max} \text{ if } km \notin \mathcal{L} \tag{32k}$$

$$\max\{\, \|(p_{km}^{\mathrm{R}}, q_{km}^{\mathrm{R}})\|\,,\ \|(p_{mk}^{\mathrm{R}}, q_{mk}^{\mathrm{R}})\|\, \} \leq S_{km}^{\max} \tag{32l}$$

$$p_{km}^{\mathrm{T}} + jq_{km}^{\mathrm{T}} = S_{km}(|V_k^{\mathrm{T}}|, |V_m^{\mathrm{T}}|, \theta_k^{\mathrm{T}}, \theta_m^{\mathrm{T}}),\ km \notin \mathcal{L} \tag{32m}$$

$$p_{mk}^{\mathrm{T}} + jq_{mk}^{\mathrm{T}} = S_{mk}(|V_m^{\mathrm{T}}|, |V_k^{\mathrm{T}}|, \theta_m^{\mathrm{T}}, \theta_k^{\mathrm{T}}),\ km \notin \mathcal{L} \tag{32n}$$

$$p_{km}^{\mathrm{R}} + jq_{km}^{\mathrm{R}} = S_{km}(|V_k^{\mathrm{R}}|, |V_m^{\mathrm{R}}|, \theta_k^{\mathrm{R}}, \theta_m^{\mathrm{R}}) \tag{32o}$$

$$p_{mk}^{\mathrm{R}} + jq_{mk}^{\mathrm{R}} = S_{mk}(|V_m^{\mathrm{R}}|, |V_k^{\mathrm{R}}|, \theta_m^{\mathrm{R}}, \theta_k^{\mathrm{R}}). \tag{32p}$$

This formulation uses the following real-valued variables, where "T" indicates true and "R," reported:

1) $|V_k^{\mathrm{T}}|, \theta_k^{\mathrm{T}}, |V_k^{\mathrm{R}}|, \theta_k^{\mathrm{R}}$ $\forall$ bus $k \in \mathcal{N}$ (true and reported voltage magnitudes and angles);
2) $P_k^{d,\mathrm{T}}, Q_k^{d,\mathrm{T}}, P_k^{d,\mathrm{R}}, Q_k^{d,\mathrm{R}}$ $\forall$ bus $k \in \mathcal{A}$ (active and reactive, true and reported loads in $\mathcal{A}$);
3) P_k^g, Q_k^g $\forall$ bus $k \in \mathcal{R}$ (generation at participating buses);
4) $\forall$ line $km \in \mathcal{E}$, $p_{km}^{\mathrm{T}}, q_{km}^{\mathrm{T}}$, and also $p_{km}^{\mathrm{R}}, q_{km}^{\mathrm{R}}$ if $km \notin \mathcal{L}$ (active and reactive, true, and reported power flows);
5) Δ (net change in active power generation).

In this formulation, power flows are represented using (32m)–(32p) [see (3)]. We include voltage variables but no current variables. However, having solved the abovementioned optimization problem, the attacker reports, for each line km with both ends in $\mathcal{A}$, a current pair $I_{km}^{\mathrm{R}}, I_{mk}^{\mathrm{R}}$ computed using the formula

$$\begin{pmatrix} I_{km}^{\mathrm{R}} \\ I_{mk}^{\mathrm{R}} \end{pmatrix} = Y_{km} \begin{pmatrix} |V_k^{\mathrm{R}}| e^{j\theta_k^{\mathrm{R}}} \\ |V_m^{\mathrm{R}}| e^{j\theta_m^{\mathrm{R}}} \end{pmatrix}$$

thereby attaining current–voltage consistency. If either $k \in \partial\mathcal{A}$ or $m \in \partial\mathcal{A}$ the true and reported voltage values are identical (see Lemma 15).

Lemma 13: Consider a feasible solution to problem (32). Let H denote either T or R (i.e., true or reported). Then, the voltages $|V_k^{\mathrm{H}}| e^{j\theta_k^{\mathrm{H}}}$ for all $k \in \mathcal{N}$ yield a solution to the power flow problem where as follows.

1) Bus k has load $P_k^{d,\mathrm{H}} + jQ_k^{d,\mathrm{H}}$ for $k \in \mathcal{A}$ and $\hat{P}_k^d + j\hat{Q}_k^d$ if $k \in \mathcal{A}^C$.
2) Bus $k \in \mathcal{G}$ has generation $P_k^g + jQ_k^g$ if $g \in \mathcal{R}$ and $\hat{P}_k^g + j\hat{Q}_k^g$ if $k \in \mathcal{G} \setminus \mathcal{R}$.
3) Line km has power flow $p_{km}^{\mathrm{H}} + jq_{km}^{\mathrm{H}}$ when $\mathrm{H} = \mathrm{R}$ and also when $\mathrm{H} = \mathrm{T}$ and $km \notin \mathcal{L}$.
4) When $\mathrm{H} = \mathrm{R}$ (reported data) the solution is fully feasible, i.e., it satisfies voltage, generation, phase angle, and power flow limits.
5) When $\mathrm{H} = \mathrm{T}$ (true data), the solution satisfies voltage, generator, and phase angle limits, but only satisfies power flow limits on lines km with both $k, m \in \mathcal{A}^C \cup \partial\mathcal{A}$. The solution is also consistent with lines in $\mathcal{L}$ being cut.

Proof: Property (3) follows from constraints (32m)–(32p). Hence, (1) and (2) follow from constraints (32c)–(32f). Properties (4) and (5) follow from constraints (32i)–(32l). ■

As a corollary to (1) and (2) of Lemma 13, a feasible solution to problem (32) satisfies, exactly, power-injection consistency.

Lemma 14: Consider a feasible solution to problem (32). The solution is consistent with a secondary-response adjustment of active power generator amounting to Δ units.

Proof: Follows from constraint (32g). ■

Lemma 15: Consider a feasible solution to problem (32). Then, (a) the true and reported voltages agree on $\mathcal{A}^C \cup \partial\mathcal{A}$. Furthermore, (b) the true and reported currents on a line km are identical if $k, m \in \mathcal{A}^C \cup \partial\mathcal{A}$.

Proof: (a) Follows from constraint (32b), and (b) is a consequence of (a). ■

Corollary 16: Suppose we compute a feasible solution to problem (32) whose objective value is strictly greater than $(S_{uv}^{\max})^2$. Then, the reported solution amounts to an undetectable attack that hides an overload on line uv.

TABLE III
TRUE AND REPORTED FLOW AT ATTACKED LINES

bus k	bus m	p_{km}^{T}	q_{km}^{T}	$\|(p_{km}^{\mathrm{T}}, q_{km}^{\mathrm{T}})\|$	S_{km}^{max}
		p_{km}^{R}	q_{km}^{R}	$\|(p_{km}^{\mathrm{R}}, q_{km}^{\mathrm{R}})\|$	
1139	1137	3.36	2.66	4.29	114.00
		3.36	2.66	4.28	
1361	1141	229.01	10.49	**229.25**	114.00
		108.51	10.49	109.02	
1141	1491	13.46	2.41	13.68	114.00
		6.20	2.39	6.64	
1141	1138	209.25	4.44	**209.29**	114.00
		98.06	5.24	98.20	

Overloads shown in the bold.

A. Computational Viability

Here, we provide a numerical example involving problem (32). We consider `case2746wp` (that has 2746 buses) from the Matpower case library. The adversary attacks the set of buses $\mathcal{A} = \{1137,\ 1138,\ 1139,\ 1141,\ 1361,\ 1491\}$ with $\mathcal{A} - \partial\mathcal{A} = \{1137, 1138, 1141, 1491\}$. In this attack, the quantity Δ in (32g) equals 135.09. We also have $\mathcal{L} = \emptyset$ (no lines are cut). The set of generators participating in secondary response is $\mathcal{R} = \{17,$ $18, 55, 57, 150, 383, 803, 804, 1996\}$ with participating factors $\alpha_k = 1/9$ for all $k \in \mathcal{R}$. A small set of participating generators makes the attacker's task more challenging due to larger power flows.

Table III shows the true and reported flow for lines where the solutions differ, with a strong overload on line $(1361, 1141)$ and $(1138, 1141)$.

VIII. Conclusion

The possibility of combined physical and data attack on power grids has gained increased attention. In principle, an attack that avoids standard detection methods is possible; we can compute such attacks on large systems in seconds of CPU time. This article focuses on stochastic defense mechanisms to augment standard detection tools. Our defenses change the stochastics of system data in a way that is recognizable by the defender but difficult to anticipate by the attacker. In future work, we will investigate distributed versions of our defense mechanisms, as well as protection against more sophisticated attackers that can also employ distributed resources and continue the attack over a prolonged period of time.

References

[1] G. Andersson, *Modelling and Analysis of Electric Power Systems*. Power Systems Laboratory, ETH Zürich, Zürich, Switzerland, 2004.

[2] A. Bergen and V. Vittal, *Power Systems Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[3] J. D. Glover, M. S. Sarma, and T. J. Overbye, *Power System Analysis and Design*. Boston, MA, USA: CENGAGE Learning, 2012.

[4] J. Zhang and L. Sankar, “Physical System Consequences of Unobservable State-and-Topology Cyber-Physical Attacks,” *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 2016–2025, Jul. 2016.

[5] Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids,” in *Proc. 16th ACM Conf. Comput. Commun. Secur.*, 2009, pp. 21–32.

[6] T. T. Kim and H. V. Poor, “Strategic protection against data injection attacks on power grids,” *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun. 2011.

[7] S. Soltan and G. Zussman, “Power grid state estimation after a cyber-physical attack under the ac power flow model,” in *Proc. IEEE Power Energy Society General Meeting*, 2017, pp. 1–5.

[8] S. Soltan and G. Zussman, “EXPOSE the line failures following a cyber-physical attack on the power grid,” *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 1, pp. 451–461, Mar. 2019.

[9] S. Soltan, M. Yannakakis, and G. Zussman, “Power grid state estimation following a joint cyber and physical attack,” *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 499–512, Mar. 2018.

[10] S. Soltan, M. Yannakakis, and G. Zussman, “REACT to cyber attacks on power grids,” *IEEE Trans. Netw. Sci. Eng.*, vol. 6, no. 3, pp. 459–473, Jul.–Sep. 2019.

[11] J. Liang, L. Sankar, and O. Kosut, “Vulnerability analysis and consequences of false data injection attack on power system state estimation,” *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3864–3872, Sep. 2016.

[12] R. Deng, P. Zhuang, and H. Liang, “CCPA: Coordinated cyber-physical attacks and countermeasures in smart grid,” *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2420–2430, Sep. 2017.

[13] D. K. Molzahn and J. Wang, “Detection and characterization of intrusions to network parameter data in electric power systems,” *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3919–3928, Jul. 2019.

[14] M. Jin, J. Lavaei, and K. Johansson, “Power grid AC-Based state estimation: Vulnerability analysis against cyber attacks,” *IEEE Trans. Autom. Control*, vol. 64, no. 5, pp. 1784–1799, May 2019.

[15] O. Vukovic, K. C. Sou, G. Dan, and H. Sandberg, “Network-aware mitigation of data integrity attacks on power system state estimation,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 6, pp. 1108–1118, Jul. 2012.

[16] S. Li, Y. Ylmaz, and X. Wang, “Quickest detection of false data injection attack in wide-area smart grids,” *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, Nov. 2015.

[17] J. Kim and L. Tong, “On topology attack of a smart grid: Undetectable attacks and contermeasures,” *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 1294–1305, Jul. 2013.

[18] J. Kim, L. Tong, and R. J. Thomas, “Subspace methods for data attack on state estimation: A data driven approach,” *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1102–1114, Mar. 2015.

[19] G. Dan and H. Sandberg, “Stealth attacks and protection schemes for state estimators in power systems,” in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Oct. 2010, pp. 214–219.

[20] D. Deka, R. Baldick, and S. Vishwanath, “Data attacks on power grids: Leveraging detection,” in *Proc. IEEE Power Energy Soc. Innovative Smart Grid Technol. Conf.*, Feb. 2015, pp. 1–5.

[21] J. Valenzuela, J. Wang, and N. Bissinger, “Real-time intrusion detection in power system operations,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1052–1062, May 2013.

[22] R. Bobba, K. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. Overbye, “Detecting false data injection attacks on DC state estimation,” in *Proc. 1st Workshop Secure Control Syst.*, 2010, pp. 1–9.

[23] A. Anwar, A. N. Mahmood, and Z. Tari, “Ensuring data integrity of OPF module and energy database by detecting changes in power flow patterns in smart grids,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3299–3311, Dec. 2017.

[24] R. Moslemi, A. Mesbahi, and J. M. Velni, “A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids,” *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4930–4941, Sep. 2018.

[25] S. Mousavian, J. Valenzuela, and J. Wang, “Real-time data reassurance in electrical power systems based on artificial neural networks,” *Elect. Power Syst. Res.*, vol. 96, pp. 285–295, 2013.

[26] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[27] X. A. Liu, D. Laverty, and R. Best, “Islanding detection based on probabilistic PCA with missing values in PMU data,” in *Proc. IEEE PES Gen. Meeting Conf. Expo.*, Jul. 2014, pp. 1–6.

[28] Z. Wang, Y. Zhang, and J. Zhang, “Principal components fault location based on WAMS/PMU measure system,” in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, pp. 1–5.

[29] F. Pasqualetti, F. Drfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[30] H. Fawzi, P. Tabuada, and S. Diggavi, “Secure estimation and control for cyber-physical systems under adversarial attacks,” *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[31] J. Zhao *et al.*, “Power system dynamic state estimation: Motivations, definitions, methodologies, and future work,” *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019.

[32] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. A. Mittelstadt, “Electromechanical mode online estimation using regularized robust RLS methods,” *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1670–1680, Nov. 2008.

[33] D. Trudnowski and J. Pierre, *Signal Processing Methods for Estimating Small-Signal Dynamic Properties From Measured Responses*. Boston, MA, USA: Springer, 2009, pp. 1–36.

[34] S. Bhela, V. Kekatos, and S. Veeramachaneni, “Enhancing observability in distribution grids using smart meter data,” *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5953–5961, Nov. 2018.

[35] S. Bhela, V. Kekatos, and S. Veeramachaneni, “Smart inverter grid probing for learning loads: Part I—Identifiability analysis,” *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3527–3536, Sep. 2019.

[36] S. Bhela, V. Kekatos, and S. Veeramachaneni, “Smart Inverter grid probing for learning loads: Part II—probing injection design,” *IEEE Trans. Power Syst.*, pp. 3537–3546, 2019.

[37] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, “Smart grid data integrity attacks,” *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1244–1253, Sep. 2013.

[38] R. Deng, P. Zhuang, and H. Liang, “CCPA: Coordinated cyber-physical attacks and countermeasures in smart grid,” *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2420–2430, Sep. 2017.

[39] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, “MATPOWER: Steady-State operations, planning, and analysis tools for power systems research and education,” *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[40] Y. Tang, G. N. Stenbakken, and A. Goldstein, “Calibration of phasor measurement unit at NIST,” *IEEE Trans. Instrum. Meas.*, vol. 62, pp. 1417–1422, Jun. 2013.

[41] K. Narendra, D. Rangana, and A. Rajapakse, "Dynamic performance evaluation and testing of phasor measurement unit (PMU) as per IEEE C37.118.1 Standard," 2018. [Online]. Available: http://www.erlphase.com/downloads/papers/Dynamic_Performance_Evaluation_and_Testing_of_PMU.pdf

[42] G. Frigo, C. Narduzzi, D. Colangelo, M. Pignati, and M. Paolone, "Definition and assessment of reference values for PMU calibration in static and transient conditions," in *Proc. IEEE Int. Workshop Appl. Meas. Power Syst.*, Sep. 2016, pp. 1–6.

[43] D. Bienstock, M. Chertkov, and M. Escobar, "Learning from power system data stream: Phasor-detective approach," 2018, *arXiv:1902.03223*.

[44] D. Bienstock, "Machine learning with PMU data," in *Proc. NASPI Work Group Meeting*, Gaithersburg, MD, USA, Mar. 2017, pp. 1–20.

[45] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

Daniel Bienstock received the Ph.D. degree in operations research from MIT, in 1985.

He is currently a Liu Family Professor of Operations Research with the Columbia University, New York, NY, USA, with affiliate appointments in applied mathematics and electrical engineering. His research focuses on optimization problems with special interest in power engineering.

Mauro Escobar received the Ph.D. degree in operations research from Columbia University, New York, NY, USA, in 2019.

He is currently a Postdoctoral Associate with Ecole Polytechnique, Paris, France.